(19) **Europäisches Patentamt**
European Patent Office
Office européen des brevets

(11) **EP 2 864 933 B1**

(12) **EUROPEAN PATENT SPECIFICATION**

(45) Date of publication and mention
of the grant of the patent:
25.09.2019 Bulletin 2019/39

(51) Int Cl.:
G06K 9/00 *(2006.01)* *G06K 9/62* *(2006.01)*

(21) Application number: **13808762.2**

(22) Date of filing: **24.04.2013**

(86) International application number:
PCT/FI2013/050459

(87) International publication number:
WO 2014/001610 (03.01.2014 Gazette 2014/01)

(54) **METHOD, APPARATUS AND COMPUTER PROGRAM PRODUCT FOR HUMAN-FACE FEATURES EXTRACTION**

VERFAHREN, VORRICHTUNG UND COMPUTERPROGRAMMPRODUKT ZUR EXTRAKTION DER MERKMALE EINES MENSCHENGESICHTS

PROCÉDÉ, APPAREIL ET PRODUIT DE PROGRAMME INFORMATIQUE D'EXTRACTION DE CARACTÉRISTIQUES DE VISAGE HUMAIN

(84) Designated Contracting States:
AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO RS SE SI SK SM TR

(30) Priority: **25.06.2012 CN 201210223706**

(43) Date of publication of application:
29.04.2015 Bulletin 2015/18

(73) Proprietor: **Nokia Technologies Oy**
02610 Espoo (FI)

(72) Inventors:
• **MA, Yong**
Beijing 100176 (CN)
• **ZOU, Yan Ming**
Beijing 100006 (CN)
• **WANG, Kong Qiao**
Beijing 101117 (CN)

(74) Representative: **Berggren Oy, Tampere**
Visiokatu 1
33720 Tampere (FI)

(56) References cited:
US-A1- 2002 012 454   US-A1- 2005 201 595
US-A1- 2006 120 571   US-A1- 2009 232 365

• **LIU ET AL: "Video-based face model fitting using Adaptive Active Appearance Model", IMAGE AND VISION COMPUTING, ELSEVIER, GUILDFORD, GB, vol. 28, no. 7, 1 July 2010 (2010-07-01), pages 1162-1172, XP027014113, ISSN: 0262-8856 [retrieved on 2009-10-02]**
• **COOTES T F ET AL: "ACTIVE APPREARANCE MODELS", IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, IEEE COMPUTER SOCIETY, USA, vol. 23, no. 6, 1 June 2001 (2001-06-01), pages 681-685, XP001110809, ISSN: 0162-8828, DOI: 10.1109/34.927467**
• **CHOI, JY ET AL.: 'Face Feature Weighted Fusion Based on Fuzzy Membership Degree for Video Face Recognition' IEEE TRANSACTIONS ON SYSTEMS, MAN AND CYBERNETICS, PART B: CYBERNETICS. vol. 42, no. 4, 12 June 2012, pages 1270 - 1282, XP011490613 Retrieved from the Internet: <URL:http://ieeexplore.ieee.org/xpl/article Details.jsp?arnumber=6216445> [retrieved on 2013-08-29]**
• **AHONEN, T ET AL.: 'Face Description with Local Binary Patterns: Application to Face Recognition' IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE. vol. 28, no. 12, December 2006, pages 2037 - 2041, XP011149704 Retrieved from the Internet: <URL:http://ieeexplore.ieee.org/xpl/article Details.jsp?arnumber=1717463> [retrieved on 2013-08-29]**

EP 2 864 933 B1

(Cont. next page)

• E. R Gast ET AL: "A Framework for Real-Time Face and Facial Feature Tracking using Optical Flow Pre-estimation and Template Tracking", , 31 December 2010 (2010-12-31), XP055576700, Retrieved from the Internet: URL:https://arxiv.org/ftp/arxiv/papers/110 1/1101.0237.pdf [retrieved on 2019-04-01]

Description

FIELD OF INVENTION

[0001] The present invention generally relates to human face recognition, and more specifically, to video sequence-based human-face recognition.

BACKGROUND OF INVENTION

[0002] With the popularity of smart phones and electronic devices with camera and video recording functions, more and more camera applications and cloud computing-based services need to obtain human-face recognition of a live video for extracting facial metadata, based on video clips from the camera or in real time online. For example, to use the human-face recognition for security operation of an access operation to an electronic device.

[0003] However, a challenge exists in directly performing human-face recognition from a video with respect to human-face recognition based on a still picture, because frame blurring and low-resolution frequently occur in the video, and in such circumstances, serious recognition error is inevitable.

[0004] By far, the video sequence-based face recognition mainly has the following three approaches:

1. Image Level Fusion

[0005] Image level fusion is one directly performed on an acquired raw image. The image level fusion generally adopts a centralized fusion system to perform the fusion processing process. It is a low-level fusion, for example, a process of determining a target property by performing image processing to a blur image containing a plurality of pixels is just an image level fusion.

[0006] For human-face recognition, an image super-resolution algorithm may be specifically employed to rebuild the human-face image. The super-resolution algorithm is a technique for enhancing the resolution of an image or video, with a purpose that the resolution of the outputted image or video would be higher than any frame of any inputted image or video. Here, "enhancing the resolution" means making the existing content much clearer or that a user can see a detail that could be perceived previously. When it is relatively difficult or costly to obtain a high-quality image or video, it is quite essential to use the super-resolution algorithm. The process of rebuilding the image super-resolution may be generally performed in three steps:

(1) pre-processing, for example, de-noising, clip, etc.;
(2) alignment, and estimating a motion vector between the low-resolution sequences, and
(3) rebuilding, and fusing information of multiple frames of the low-resolution image.

[0007] The super-resolution rebuilding process of an image generally needs three-dimensional modeling, which results in a cumbersome computational complexity.

[0008] Besides, there is also a scheme of de-blurring an image and then restoring the image specifically directed to the cause of the blur, for example, restoring for a motion blur, and restoring for a defocus, etc. Its main purpose is to generate a clear picture so as to perform works such as recognition and judgment.

[0009] However, currently, the image level fusion for human-face recognition is mainly used for visual inspection, which is not very flexible and is quite sensitive to environment (for example noise), misalignment, and the like.

2. Feature level fusion

[0010] The feature level fusion mainly extracts a local feature of a human face from among each frame of a video image. Since the same kind of samples have a certain distribution in space, image set sub-space (mutual sub-space) and manifold learning may be employed to reduce dimensions of the feature space of the sample, and then the dimension-reduced sample feature space is matched to the logged sample egien space, thereby performing human-face recognition.

[0011] In this scheme, all local features in a feature vector of the human face come from the same frame; therefore, it does not break away from the constraint of frame.

3. Classifier level fusion

[0012] The classifier level fusion builds a multi-scale target classifier and a pose determiner, respectively, mainly based on scale variation, pose variation, and image feature information of an object, and estimates a confidence level of a target recognition result, a weight of adjacent frame pose variation, and a target scale weight; the recognition result of each frame is compared to an image sample in a database to score each frame. And then, the target image fusion is performed based on the score of each frame. However, the classifier level recognition mainly relies on the classifier of a single frame, and the decision based on the score of each frame; thus, it still has a drawback of inaccurate classification caused by insufficient feature extraction; besides, in a complex dynamic environment, there are fewer appropriate classifier algorithms to implement recognition of a dynamic target.

Non-Patent Literature 1: "A Framework for Real-Time Face and Facial Feature Tracking using Optical Flow Pre-estimation and Template Tracking", E. R Gast, April 2012;
Non-Patent Literature 2: "Video-based face model fitting using Adaptive Active Appearance Model", X.

Liu, July 2010

[0013] Therefore, there is a need for rapidly, accurately and robustly recognizing a human-face image from a video sequence.

SUMMARY OF INVENTION

[0014] The present invention resolves the above and other problems by providing a novel method for extracting a human-face local feature. According to the present invention, during a human-face tracking process of a video sequence, a set of face images belonging to a same person is obtained, and correspondence relationships between respective facial points of the face image are obtained using a facial point tracking technology. Human-face local feature vector components are extracted from respective facial points of the face image, and the extracted human-face local feature vector components are fused to obtain human-face local feature vectors base on the analysis of the face image. Next, the obtained human-face local feature vectors are combined into a human-face global feature vector of the entire human face so as to perform human-face recognition.

[0015] According to one aspect of the present invention, there is provided a method comprising the features of the independent claim 1.

[0016] According to another aspect of the present invention, there is provided an apparatus for obtaining a human-face feature vector from a video sequence, as defined in the independent claim 9.

[0017] According to a further aspect of the present invention, there is provided a computer program product as defined in the independent claim 10.

[0018] The embodiment of the invention relates to a method as claimed in claim 1 and corresponding apparatus as claimed in claim 9 and a corresponding computer program code as claimed in claim 10. Other references to "embodiments" are to be understood as examples which are useful for understanding the invention.

[0019] According to the present invention, one or more human-face local feature vector components as fused may come from human-face local feature vector components of different frames. However, in the aforementioned feature level fusion method, all human-face local feature vector components in one human-face local feature vector come from the same frame, which does not break away from the constraints of frame.

[0020] Besides, according to the present invention, a detected human-face area will be divided under different scales into a plurality of local patches with a predetermined size, and human-face local feature vector components are extracted from these local patches having a mutual correspondence relationship, respectively, and are combined together to form a human-face local feature vector representing a facial point corresponding to each local patch, and then the human-face local feature vectors representing the facial points corresponding to re-

spective local patches and obtained at each local patch size are combined to form a human-face global feature vector describing the entire human face so as to be used for human-face recognition. Besides, according to the present invention, all human-face global feature vectors obtained under different scales may be further combined to form a human-face global feature vector set so as to perform human-face recognition. All human-face global feature vectors obtained under different scales are combined to form a human-face global feature vector set, and the finally obtained multiple human face global feature vectors are more robust to pose offset of the human face and environment influence.

BRIEF DESCRIPTION OF SEVERAL VIEWS OF DRAWINGS

[0021] The present invention has been generally described above. Now, the present invention will be described with reference to the accompanying drawings that are not necessarily drawn by scale, wherein:

Fig. 1 shows a schematic diagram of a basic principle of the present invention.
Fig. 2 shows a flowchart of a method for generating human-face local feature vectors according to an exemplary embodiment of the present invention;
Fig. 3 shows a schematic diagram of a correspondence relationship between respective patches of a human-face area in adjacent frames of a video sequence according to an exemplary embodiment of the present invention;
Fig. 4 shows a schematic diagram of dividing a human-face area into local patches of different sizes according to an exemplary embodiment of the present invention;
Fig. 5 shows a block diagram of an apparatus configured to generate human-face local feature vectors according to an exemplary embodiment of the present invention; and
Fig. 6 shows a schematic diagram of an apparatus for generating human-face local feature vectors according to an exemplary embodiment of the present invention.

[0022] In the drawings, same or corresponding reference signs indicate the same or corresponding parts. Moreover, the number of components, members, and elements as illustrated in the drawings are only for exemplary illustration, not for limitation.

DETAILED DESCRIPTION OF INVENTION

[0023] With reference to Fig. 1, it generally illustrates the basic principle of the present invention. The four images from left to right represent four frames (from frame i to frame i+3) of images in a video sequence, which shows change of a person from far to near with respect

to a camera lens. According to the present invention, first, human-face area detection is performed with respect to each frame of image. Then, the detected human-face area is divided into a plurality of local patches with a predetermined size, as shown by the grid in each human-face area in Fig. 1. Next, the correspondence relationship between each patch of a plurality of human-face areas may be determined by optical flow. Specifically, it may be implemented by KLT, and those skilled in the art may know the method of obtaining the correspondence relationship between the same facial point in immediately adjacent two frames in a video sequence by using KLT for example from "Good features to track" by Jianbo Shi and Carlo Tomasi, Proc. IEEE Comput. Soc. Conf. Comput. Vision and Pattern Recogn., pages 593-600, 1994. With reference to Fig. 1, human-face local feature vector components are extracted from respective patches of the human-face area of the ith frame of image to form human-face local feature vectors representing the facial points corresponding to the local patches. Next, human-face local feature vector components are extracted sequentially from local patches in a subsequent image frames having a correspondence relationship with the local patch of the ith frame of image and are evaluated (for their occluded level, etc.), so as to update the human-face local feature vectors representing the facial points corresponding to the local patches using the human-face local feature vector components extracted from the local patches in the subsequent image frames having a corresponding relationship with the local patches of the ith frame of image. Finally, the obtained human-face local feature vectors are combined into a human-face global feature vector describing the whole human face, and a template-based facial recognition method, such as principle component analysis (PCA)+ linear discrimination analysis (LDA), is employed to compare with the human-face feature vectors in a human-face gallery to perform human-face recognition. Those skilled in the art may know the above method for human-face recognition for example from W. Zhao, R. Chellappa, A. Krishnaswamy, Discriminant Analysis of Principal Components for Face Recognition, Proc. of the 3rd IEEE International Conference on Face and Gesture Recognition, FG'98, 14-16 April 1998, Nara, Japan, pp. 336-341 and A.M. Martinez, A.C. Kak, PCA versus LDA, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 23, No. 2, 2001, pp. 228-233.

[0024]   Fig. 2 shows a flow chart of a method according to an exemplary embodiment of the present invention. It should be understood that each block, step or operation in the flow chart, and/or a combination of blocks, steps or operations in the flow chart may be implemented by various means. The means for implement blocks, steps or operations in the flow chart and/or combinations of the blocks, steps, or operations in the flow chart may comprise hardware, firmware and/or software, the software comprising one or more computer program code instructions, program instructions, or executable computer

readable program code instructions. In one exemplary embodiment, one or more processes described here may be implemented by program code instructions. At this point, the program code instructions implementing the process described here may be stored by a memory device of the apparatus or stored thereon and executed by a processor. It should be noted that any such program code instructions may be loaded onto a computer or other programmable device to produce a machine, such that the instructions executed on the computer or other programmable device to create means for implementing the functions designated in the blocks, steps, or operations of the flow chart. These program code instructions may be stored in a computer readable storage medium, which may guide a computer, processor or other programmable device to work in a particular manner, for producing an article according to the instructions stored in the computer readable storage medium, the instructions comprising instruction means for implementing the functions designated in the blocks, steps or operations of the flow chart. The program code instructions may also be loaded on the computer, processor or other programmable devices, such that a series of operation steps may be performed on the computer, processor, or other programmable device or performed thereby, to thereby produce a process implemented by the computer, such that the instruction executed on the computer, processor or other programmable device provide steps for implementing the functions designated in the blocks, steps or operations of the flow chart.

[0025]   Therefore, the blocks, steps or operations in the flow chart support a combination of means for performing designated functions, a combination of steps for performing designated functions, and program code instruction means for performing designated functions. It would be further appreciated that one or more blocks, steps or operations in the flow chart, and a combination of blocks, steps or operations in the flow chart may be implemented by a dedicated hardware-based computer system or a dedicated hardware and program code instructions, wherein the computer system performs designated functions or steps.

[0026]   With reference to Fig. 2, in step S210, a same human-face image in a plurality of image frames of a video sequence is detected. The video sequence may have a plurality of frames having an image of the same human face, and of course, in each frame of image, there may be other images besides the human face. In order to implement the method of the present invention, it is first required to detect the same human-face image in the plurality of frames of images, which may be implemented by a human-face detector. As an example, a human-face detector may be obtained through the following operation: during an offline training phase, millions of or more normalized (for example 20*20 pixels) human-face image data and non-human-face image data are collected. Based on these normalized training data, one or more LBP features may be extracted around each pixel, and

then these features extracted from all pixels may be used to form a LBP feature pool. Based on each LBP feature, a decision tree is used to build a weak classifier. By using an AdaBoosting learning strategy, a best weak classifier is selected in each iteration. Finally, based on a weighted combination of many weak classifiers, a strong classifier having a high human-face detection rate and a low mis-alarm rate is formed. This final strong classifier may be used as the human-face detector of the present invention. The human-face detector may detect a rough position, scale, roll angle, and pitch angle of the human face.

[0027] In step S220, the detected human-face image is divided into a plurality of local patches with a predetermined size, wherein each local patch is around or near a human-face feature point.

[0028] The position of a human-face feature point (for example, the center of an eye, a corner of an eye, a corner of the mouth, and the like) may be precisely located on the detected face image by a facial point locator. According to one embodiment of the present invention, a local binary pattern (LBP) feature + AdaBoosting classifier is employed to perform human-face image detection and facial point location. Those skilled in the art may know the human-face image detection and facial point location method for example from Ojala & Maenpaa (2002) Multiresolution gray-scale and rotation invariant texture classification with Local binary Patterns, IEEE PAMI 24(7):971-987. The facial point locator may be built in a manner similar to the human- face detector by using a corresponding training sample. Specifically, each time when the human- face area is scaled, all sub-windows of training data with the same resolution are evaluated through a trained classifier, and all positive responses are fused together based on the positions and confidence levels of the sub-windows so as to output the final human-face detection result, thereby implementing the facial point location process.

[0029] Specifically, the step S220 may be implemented through the following steps: after detecting the same human-face image in a plurality of image frames of the video sequence and accomplishing location of a major facial point, the facial area is cut and normalized and scaled into a predetermined resolution; at this point, the grids may be placed on the human-face area, as shown in Fig. 1 and Fig. 4 (Fig. 1 shows a 5*5 grid, and Fig. 4 shows a 6*6 grid, 8*8 grid, and 16*16 grid), wherein each grid area (or referred to as a local patch) is around or near a facial point.

[0030] Next, in step S230, a correspondence relationship between each local patch of the same human-face image in the plurality of image frames of the video sequence is determined.

[0031] With reference to Fig. 3, it shows a correspondence relationship between respective local patches of human-face areas of adjacent frames of a video sequence. Fig. 3 shows the ith frame of image and the following i+1st frame of image in the video sequence. Although Fig. 3 merely shows two adjacent frames of im-

ages, the method of the present invention is not merely limited to two frames of images, but may be applied to a consecutive plurality of frames of images.

[0032] According to the embodiments of the present invention, as depicted above with reference to Fig. 1, a method such as optical flow field may be used to obtain a correspondence relationship between the same facial point (i.e., each patch) in immediately adjacent frames.

[0033] After the operation of step S230 is accomplished, the process enters into step S240. Here, human-face local feature vector components extracted from respective local patches having a mutual correspondence relationship are used to form human-face local feature vectors representing facial points corresponding to the local patches. Specifically, it may also be implemented through the following steps: first, human-face local feature vector components are extracted from respective patches having a mutual correspondence relationship in different image frames. Then, the pose of each human face in different image frames is determined, which may be implemented based on a LBP predicted pose angle by extracting LBP feature and using Canonical Correlation Analysis recursion. After determining the pose of each human face in each image frame, the human-face local feature vector components extracted from respective patches are used to form human-face local feature vectors representing facial points corresponding to the local patches, which may be expressed in the following equation:

$$V_P = \{V_{i,P}, V_{i+1,P}, \cdots, V_{i+n,P}\},$$

Wherein V_p denotes the human-face local feature vector of the facial point corresponding to each patch P, and $V_{i,P}$ denotes a human-face local feature vector component extracted from the patch P of the ith frame.

[0034] According to one embodiment of the present invention, the step S240 may specifically comprise the following steps: identifying different poses of the same human face in a plurality of image frames of the video sequence; extracting human-face local feature vector components merely from respective un-occluded local patches having a mutual correspondence relationship based on the identified different poses of the same human face in the plurality of image frames; combining the human-face local feature vector components extracted from the respective un-occluded local patches having a mutual correspondence relationship to form human-face local feature vectors representing facial points corresponding to the local patches.

[0035] Specifically, referring to Fig. 3, through determining the human-face poses, it may be determined that the human face in the ith frame of image looks to the right side, while the human face in the i+1st frame of image looks substantially to the forward. At this point, it may be determined that in the ith frame of image, the right-side

area of the human face is occluded. Therefore, it may be considered to use the human-face local feature vector components extracted from the corresponding patches in the right-side area of the i+1st frame of the image as a part of the human-face feature vectors representing the facial points corresponding to the patches, instead of the human-face local feature vector components extracted from the corresponding patches in the right-side area of the ith frame of image, because at this point, the human-face right-side area in the ith frame of image is occluded. Specifically, $P_{j,k}$ here denotes each patch resulting from dividing the human-face area in the ith frame of image, wherein j denotes the row number where the patch is located (from top to down), while k denotes the column number where the patch is located (from left to right). Correspondingly, $P'_{j,k}$ denotes each patch in the i+1 frame having a corresponding relationship with each $P_{j,k}$. For example, for each patch $P_{1,1}$ in the ith frame of image, it may be determined through judging the human-face pose that the human-face area corresponding to the patch might be occluded; thus, at this point, the human-face local feature vector component is extracted from the patch $P'_{1,1}$ in the i+1 frame of image having a correspondence relationship therewith as a part of the human-face local feature vector of the facial point corresponding to the patch. Similarly, it is also suitable for the patches $P_{2,1}$, $P_{3,1}$, $P_{1,2}$, $P_{1,3}$ and the like which are likewise occluded. Although Fig. 3 merely illustrates two consecutive frames of images, it is apparent that for subsequent i+2, i+3, ... images, if any one of the patches $P''_{1,1}$, $P'''_{1,1}$ ... having a correspondence relationship with $P_{1,1}$ is judged as being occluded, no human-face local feature vector components are extracted from this patch. On the contrary, if it is judged that none of these patches $P''_{1,1}$, $P'''_{1,1}$ ... is occluded, the human-face local feature vector component may be extracted from the patch as a part of the feature vector set corresponding to the facial point.

[0036] An advantage of this embodiment of the present invention lies in that since the step of extracting human-face local feature vector components from the occluded local patch is discarded, it avoids the over large background noise caused by the human-face local feature vector components extracted from the occluded local patch to the final human-face local feature vector, which would affect the accuracy of human-face recognition. Besides, since no human-face local feature vector components are extracted from the occluded local patches, it reduces the work load of extracting human-face feature local vector components and shortens the time for extracting human-face local feature vector components.

[0037] According to another embodiment of the present invention, the step S240 may specifically comprise the following steps: identifying different poses of the same human face in a plurality of image frames of the video sequence; weight combining, based on the identified different poses of the same human face in different image frames the human-face local feature vector components extracted from respective local patches

having a mutual corresponding relationship to form human-face local feature vectors representing the facial points corresponding to the local patches.

[0038] Specifically, further referring to Fig. 3, at this point, although it has been judged that the human-face area corresponding to the patch $P_{1,1}$ in the ith frame image is occluded, a human-face local feature vector component is still extracted from the patch. Likewise, a human-face local feature vector component is also extracted from the patch $P'_{1,1}$ having a correspondence relationship with $P_{1,1}$ in the i+1st fame of image. Since it has been judged at this point that the patch $P_{1,1}$ of the ith frame of image is occluded, when forming a human-face local feature vector representing the facial point corresponding to the patch, it will be considered to assign the human-face local feature vector a relatively small weight such that it will not cause an over large background noise component to the finally formed human-face local feature vector; while for the patch $P'_{1,1}$ in the i+1st frame, since it has been judged that the i+1st frame of image is substantively facing forward, a relatively large weight may be assigned thereto, because it will not cause a very large background noise component to the formed local feature vector. Likewise, although Fig. 3 merely illustrates two consecutive frames of images, it is apparent that for subsequent i+2, i+3, ... images, the same method is also applicable. At this point, the obtained human-face local feature vector may be expressed as:

$$V_P = \{a_1 V_{i,P}, a_2 V_{i+1,P}, \cdots, a_n V_{i+n,P}\},$$

wherein a_1, a_2, ..., a_n denote weights of the human-face local feature vector components extracted from the patches having a correspondence relationship in each frame, respectively.

[0039] The advantage of another embodiment of the present invention lies in that through extracting human-face local feature vector components from each local patch having a mutual correspondence relationship and weight combining them to form a human-face local feature vector, the resulting human-face local feature vector comprises more human-face local feature vector components than a human-face local feature vector formed by human-face local feature vector components merely extracted from un-occluded local patches, which inevitably improves accuracy for subsequent human-face recognition. Meanwhile, by assigning different weights to the human-face local feature vector components extracted from each local patch having a relevant correspondence relationship based on human face poses, the human-face local feature vector components extracted from each local patch would play different roles in the subsequent human-face recognition process, which, on the one hand, restricts the over large background noise caused by the human-face local feature vector components extracted from the occluded local patches to the

formed human-face local feature vector, and further improves the accuracy of human-face recognition.

[0040] According to the embodiments of the present invention, there further comprises a step of resizing the plurality of local patches. Specifically, it may be realized through re-sizing the grids on the human-face area, as shown in Fig. 4.

[0041] Fig. 4 shows a diagram of dividing a human-face area into local patches of different sizes according to an embodiment of the present invention. Fig. 4 shows the same frame of image, wherein the first image (from left to right) shows a standard 64*64-pixel human-face image, which is obtained after detecting and locating the human-face area in the image as described above. The second image shows dividing the human-face image in the first image into a 6*6-grid patches, the third image shows dividing the human-face image in the first image into a 8*8-grid patches, and the last image shows dividing the human-face image in the first image into a 16*16-grid patches. In these images, each patch in the grid patches is around or near a facial point.

[0042] After resizing the plurality of local patches, steps S230-S240 are repeated to extract human-face local feature vector components from each resized local patches having a mutual corresponding relationships to form a human-face local feature vector representing the facial point corresponding to the resized local patch.

[0043] Specifically, with reference to Fig. 4, based on dividing different scales of a human-face area as shown in the second to fourth images, the human-face local feature vector forming process is unanimously executed with reference to steps S230-240. Then, all resulting human-face local features are combined together to form a human-face global feature vector describing the whole human face. Next, human-face recognition is performed using the resulting human-face global feature vector.

[0044] The advantage of this embodiment lies in that in specific operation, for example, the located human-face area may be first divided into a 6*6 grid, and at this point, human-face local feature vectors are obtained by extracting human-face local feature vector components using the method as depicted with reference to steps S230-S240, and then, the resulting human-face local feature vectors are combined into a human-face global feature vector which is compared with human-face feature vectors in a human face gallery to perform human-face recognition. If the human face can be recognized now, the subsequent step of further dividing the human-face area will become unnecessary. In this way, it may save computational resources and improve the recognition speed, because compared with dividing the human-face area into 8*8 grid, dividing the human-face area into 6*6 grid apparently may reduce the work load for extracting the human-face local feature vector components corresponding to the patches.

[0045] Although the speed of extracting human-face local feature vector components can be improved by dividing the human-face areas into less grid patches to extract human-face local feature vector components, it also brings about the problem of inaccurate recognition. It is because a larger grid may be very sensitive to facial point change. Therefore, according to the present invention, after dividing the human-face area into 6*6 grid patches to extract the human-face local feature vector components, the human-face area is further divided into 8*8 grid patches (the second image), 16*16 grid patches (the third image) ... to continue extracting human-face local feature vector components to obtain corresponding human-face local feature vectors, and then the human-face local feature vectors obtained respectively under different divisions are combined into human global feature vectors describing the whole human-face. Finally, all human-face global feature vectors obtained under different divisions are combined together to form a human-face global feature vector set for subsequent human-face recognition operation.

[0046] The practice shows that the human-face local feature vector components extracted from the patches divided in different scales in the human-face area have been proved to be robust to human-face poses and facial expression variations (for this point, please refer to "SIFT feature matching for face recognition" in BMVC 2009).

[0047] According to one embodiment of the present invention, after step S240, there may further comprise a step of combining the resulting human-face local feature vectors representing facial points corresponding to respective local patches to form a human-face global feature vector. The resulting human-face global feature vector therefore represents the human-face local feature vectors representing all facial points corresponding to all local patches, and then, the resulting human-face global feature vector may be used to compare with human-face feature vectors in the human-face gallery to perform human-face recognition. Similarly, after resizing each local patch, the resulting human-face local feature vectors representing the facial points corresponding to each resized local patches may further be combined to form a human-face global feature vector, and the resulting human-face global feature vector therefrom comprises human-face local feature vectors representing all facial points corresponding to all resized local patches, and then, the resulting human-face global feature vector may be used to compare with the human-face feature vectors in the human face gallary to perform human face recognition.

[0048] According to another embodiment of the present invention, there further comprises a step of combining human-face global feature vectors obtained under different local patch sizes to form a human-face global feature vector set. The human-face global feature vector set resulting from this step comprises a plurality of human-face global feature vectors, wherein each human-face global feature vector comprises human-face local feature vector representing all facial points corresponding to local patches of a certain size. Since the human-face global feature vector set resulting from this operation comprises a plurality of human-face global feature vec-

tors formed by combining human-face local feature vectors extracted from different sizes of local patches having a mutual correspondence relationship, it is robust to change in human face poses and human expressions; therefore, it may play a better role during the process of comparing with the human-face feature vectors in the human-face gallary to perform human face recognition.

[0049] Fig. 5 shows an exemplary device 500 that may provide a correlation indication. According to various exemplary embodiments of the present invention, the device 500 may be implemented as a wired or wireless communication device of any type or included therein as a component of these devices. The wired or wireless communication device is for example a mobile terminal, an access point such as a base station, or any combination thereof, wherein the mobile terminal includes for example a mobile phone, a portable digital assistant (PDA), a pager, a mobile TV, a gaming device, a mobile computer, a laptop, a camera, a recorder, an audio/ video player, a radio receiver and/or a global positioning system (GPS) device, an access point such as a base station, or any combination of the foregoing devices, and the like. Besides, the device 500 may be configured to implement various aspects of the present invention as depicted here, for example, including various exemplary methods of the present invention, wherein these methods may be implemented by a hardware or software configured processor, a computer readable medium, and the like.

[0050] The device 500 may comprises a processor 505, a memory device 510, and a communication interface 520, or may communicate with these components. In some embodiments, the device 500 may also comprise a user interface 515. The processor 505 may be implemented into various means including for example, a microprocessor, a co-processor, a controller, or various kinds of other processors, including an integrated circuit, for example, an ASIC (application-specific integrated circuit), FPGA (field programmable gate array) or hardware accelerator. In an exemplary embodiment, the processor 505 may be configured to execute the instructions stored in the memory device 510 or the instructions accessible to the processor 505. The processor 505 may also be configured to facilitate communication via the communication interface 520 by for example controlling the hardware and/or software in the communication interface 520.

[0051] The memory device 510 may be a computer readable memory medium which may include a volatile and/or non-volatile memory. For example, the memory device 510 may comprise a random access memory (RAM) including a dynamic and/or static RAM, an on-chip or off-chip cache memory, etc. Besides, the memory device 510 may comprise a non-volatile memory which may be embedded and/or movable, and for example may comprise a read-only memory, a flash memory, a magnetic storage device (for example, a hard disk, a floppy driver, a magnetic tape, and the like), an optical disk driver and/or medium, a non-volatile random access memory (NVRAM), and the like. The memory device 510 may comprise a cache area for temporarily storing data. In this point, some or all memory devices 510 may be included in the processor 505.

[0052] Besides, the memory device 510 may be configured to store memory information, data, application, computer-readable program code instructions, and the like, such that the processor 505 and the device 500 can execute various functions according to the exemplary embodiments of the present invention. For example, the memory device 510 may be configured to buffer the input data processed by the processor 505. Additionally or alternatively, the memory device 510 may be configured to store the instructions executed by the processor 505.

[0053] The communication interface 520 may be any device or apparatus implemented by hardware, software, or a combination of hardware and software; these devices or apparatuses are configured as any other device or module receiving and/or sending data from/to a network and/or communicating with the device 500. In this point, the communication interface 520 may for example comprise an antenna, a transmitter, a receiver, a transceiver, and/or support hardware including a processor or software for supporting communication with the network 525 which may be any type of wired or wireless network. Via the communication interface 520 and the network 525, the device 500 may communicate with various other network entities. In this point, the network 525 may comprise an access point.

[0054] The communication interface 520 may be configured to provide communication according to any wired or wireless communication standard. For example, the communication interface 520 may be configured for communication according to the following contents: second-generation (2G) wireless communication protocol IS-136 (time-division multi-access (TDMA)), GSM (global mobile communications system), IS-95 (code-division multi-access (CDMA)), third-generation (3G) wireless communication protocol, such as universal mobile telecommunications system (UMTS), CDMA2000, wideband CDMA (WCDMA) and time-division synchronous CDMA (TD-SCDMA), 3.9-generation (3.9G) wireless communication protocol, for example, evolved universal terrestrial wireless access network (E-UTRAN), and the fourth-generation (4G) wireless communication protocol, advanced international mobile telecommunications (IMT-Advanced) protocol, including LTE-advanced long-term evolution (LTE) protocol, etc. Besides, the communication interface 520 may be configured to provide communication according to the following technologies, such as radio frequency (RF), infrared (IrDA), or any one of a plurality of different wireless networking technologies, including WLAN technologies, such as IEEE 802.11 (for example, 802.11a, 802.11b, 802.11g, 802.11n, etc.), wireless local area network (WLAN) protocol, such as WorldWide Interoperability for Microwave Access (WiMAX) of IEEE 802.16 and/or IEEE 802.15, Bluetooth (BT), ultra wideband (UWB) wireless personal area network (WPAN) technologies, and the like.

[0055] The user interface 515 may communicate with the processor 505 to receive user input at the user interface 515 and/or provide output to the user, for example, audible, visual, mechanical or other output indication. The user interface 515 for example may include a keyboard, a mouse, a joy stick, a display (for example, a touch screen display), a microphone, a speaker, or other input/output mechanism. In some exemplary embodiments, such as when the device 500 is implemented as a server, the user interface may be limited or even eliminated.

[0056] The human-face detector 530, facial point locator 532, human-face tracker 534, and human-face local feature vector generator 536 of the device 500 may be any apparatus or device implemented by hardware, software, or a combination thereof, such as a processor 505 for implementing software instructions or hardware configured processor 505 which is configured to perform the functions of the human-face detector 530, facial point locator 532, human-face tracker 534, and human-face local feature vector generator 536 as depicted here. In an exemplary embodiment, the processor 505 may include or control the human-face detector 530, the facial point locator 532, the human-face tracker 534, and the human-face local feature vector generator 536. In various exemplary embodiments, the human-face detector 530, the facial point locator 532, the human-face tracker 534, and the human-face local feature vector generator 536 may reside on different devices, such that some or all functions of the human-face detector 530, the facial point locator 532, the human-face tracker 534, and the human-face local feature vector generator 536 may be executed by the first device, and the remaining functions of the human-face detector 530, the facial point locator 532, the human-face tracker 534, and the human-face local feature vector generator 536 may be executed by one or more other devices.

[0057] The human-face detector 530 of the device 500 is configured to detect a same human-face image in a plurality of image frames of a video sequence. The video sequence may be a video sequence stored in the memory device 510 or a video sequence received from an external input device or external memory via the user interface 515. The human-face detector 530 detects the human-face image in a plurality of image frames of the video sequence in a manner as depicted in step S210 with reference to Fig. 2.

[0058] The facial point locator 532 of the device 500 is configured to divide the detected human-face image into a plurality of local patches with a predetermined size in a manner as depicted in step S220 with reference to Fig. 2, wherein each local patch is around or near a human face feature point. Specifically, it may be implemented by adding grid on the detected human-face image.

[0059] The human-face tracker 534 of the device 500 is configured to determine a correspondence relationship between respective local patches of the same human-face image in a plurality of image frames of the video

sequence, and the human-face local feature vector generator 536 of the device 500 is configured to use human-face local feature vector components extracted from respective local patches having a mutual correspondence relationship to form human-face local feature vectors representing the facial points corresponding to the local patches, as depicted in step S230 of Fig. 2.

[0060] According to one embodiment of the present invention, the human-face local feature vector generator 536 is further configured to identify different poses of the same human face in a plurality of image frames of the video sequence; extract human-face local feature vector components merely from respective un-occluded local patches having a mutual correspondence relationship based on the identified different poses of the same human face in the plurality of image frames; combine the human-face local feature vector components extracted from the respective un-occluded local patches having a mutual correspondence relationship to form human-face local feature vectors representing the facial points corresponding to the local patches.

[0061] According to another embodiment of the present invention, the human-face local feature vector generator 536 is further configured to identify different poses of the same human face in a plurality of image frames of the video sequence; and weight combine, based on the identified different poses of the same human face in different image frames, the human-face local feature vector components extracted from respective local patches having a mutual corresponding relationship to form human-face local feature vectors representing the facial points corresponding to the local patches.

[0062] According to the embodiments of the present invention, the facial point locator 534 is further configured to resize the plurality of local patches to obtain the human-face local feature vectors representing the facial points corresponding to each resized local patches.

[0063] According to the embodiments of the present invention, the human-face local feature vector generator 536 is further configured to combine the resulting human-face local feature vectors representing the facial points corresponding to respective local patches to form a human-face global feature vector.

[0064] According to the embodiments of the present invention, the human-face local feature vector generator 536 is further configured to combine the resulting human-face local feature vectors representing the facial points corresponding to respective resized local patches to form a human-face global feature vector.

[0065] According to the embodiments of the present invention, the human-face local feature vector generator 536 is further configured to combine the resulting human-face global feature vectors obtained under different local patch sizes to form a human-face global feature vector set.

[0066] According to the embodiments of the present invention, the human-face local feature vector generator 536 is further configured to use the resulting human-face

global feature vectors to perform human-face recognition.

[0067] According to the embodiments of the present invention, the human-face local feature vector generator 536 is further configured to use the resulting human-face global feature vector set to perform human-face recognition.

[0068] Fig. 6 shows a diagram of an apparatus 600 for generating a human-face local feature vector according to an exemplary embodiment of the present invention.

[0069] Specifically, the apparatus for generating a human-face local feature vector according to the present invention comprises: means 610 for detecting a same human-face image in a plurality of image frames of a video sequence; means 620 for dividing the detected human-face image into a plurality of local patches of a predetermined size, wherein each local patch is around or near a human-face feature point; means 630 for determining a correspondence relationship between respective local patches of the same human-face image in the plurality of image frames of the video sequence; and means 640 for using human-face local feature vector components extracted from respective local patches having a mutual correspondence relationship to form human-face local feature vectors representing the facial points corresponding to the local patches.

[0070] According to one embodiment of the present invention, wherein the means 640 for using human-face local feature vector components extracted from respective local patches having a mutual correspondence relationship to form the human-face local feature vectors representing facial points corresponding to the local patches further comprises: means for identifying different poses of the same human face in a plurality of image frames of the video sequence; means for extracting human-face local feature vector components only from respective un-occluded local patches having a mutual correspondence relationship based on the identified different poses of the same human face in the plurality of image frames; and means for combining the human-face local feature vector components extracted from the respective un-occluded local patches having a mutual correspondence relationship to form human-face local feature vectors representing the facial points corresponding to the local patches.

[0071] According to another embodiment of the present invention, wherein the means 640 for using human-face local feature vector components extracted from respective local patches having a mutual correspondence relationship to form human-face local feature vectors representing facial points corresponding to the local patches further comprises: means for identifying different poses of the same human face in a plurality of image frames of the video sequence; and means for weight combining the human-face local feature vector components extracted from respective local patches having a mutual correspondence relationship into human-face local feature vectors representing the facial points corresponding to the local patches based on the

identified different poses of the same human face in the different image frames.

[0072] According to the embodiments of the present invention, dividing the detected human-face image into a plurality of local patches having a predetermined size is implemented by adding a grid on the detected human-face image.

[0073] According to the embodiments of the present invention, the apparatus further comprises means for resizing the plurality of local patches to obtain human-face local feature vectors representing the facial points corresponding to respective resized local patches.

[0074] According to the embodiments of the present invention, the apparatus further comprises a means for combining the resulting human-face local feature vectors representing facial points corresponding to respective local patches to form human-face global feature vectors.

[0075] According to the embodiments of the present invention, the apparatus further comprises a means for combining the resulting human-face local feature vectors representing facial points corresponding to respective resized local patches to form human-face global feature vectors.

[0076] According to the embodiments of the present invention, the apparatus further comprises means for combining human-face global feature vectors obtained under different local patch sizes to form a human-face global feature vector set.

[0077] According to the embodiments of the present invention, the apparatus further comprises means for recognizing a human face using the resulting human-face global feature vectors.

[0078] According to the embodiments of the present invention, the apparatus further comprises means for recognizing a human face using the resulting human-face global feature vector set.

[0079] Those skilled in the art benefited from the teaching embodied in the above depictions and the associated drawings would contemplate various modifications and other embodiments of the present invention as depicted here. Besides, although the above depiction and associated drawings have described exemplary embodiments in some exemplary combination environments of the elements and/or functions, it should be understood that different combinations of elements and/or functions may be provided by alternative embodiments.

Claims

1. A method comprising:

    obtaining a video sequence comprising a set of face images belonging to a same person;
    detecting (S210) human-face images in adjacent frames of a plurality of image frames of the video sequence;
    dividing (S220) the detected human-face imag-

es into a plurality of local patches with a predetermined size, wherein each local patch corresponds to a human-face area around a human-face feature point;

determining (S230), using optical flow field, a correspondence relationship between respective human-face feature points of the human-face images in the adjacent frames of the plurality of image frames of the video sequence; identifying different poses of the human-face in a plurality of image frames of a video sequence; determining if any human-face area is occluded based on the identified different poses of the human-face in the plurality of image frames; **characterized in that** the method further comprises extracting human-face local feature vector components only from local patches corresponding to un-occluded human-face areas having a mutual correspondence relationship and discarding extracting human-face local feature vector components from local patches corresponding to occluded human-face areas; and

combining the human-face local feature vector components extracted from local patches in adjacent frames corresponding to un-occluded human-face areas having a mutual correspondence relationship to form a human-face local feature vector representing the facial points corresponding to the local patches.

2. The method according to claim 1, wherein dividing the detected human-face image into a plurality of local patches with a predetermined size is implemented by adding a grid on the detected human-face image.

3. The method according to claim 1 or 2, further comprising resizing the plurality of local patches to obtain the human-face local feature vectors representing the facial points corresponding to respective resized local patches.

4. The method according to claim 3, further comprising combining the resulting human-face local feature vectors representing the facial points corresponding to respective local patches to form human-face global feature vectors.

5. The method according to claim 4, further comprising combining the resulting human-face local feature vectors representing the facial points corresponding to respective resized local patches to form human-face global feature vectors.

6. The method according to claim 5, further comprising combining human-face global feature vectors obtained under different local patch sizes to form a human-face global feature vector set.

7. The method according to claim 4 or 5, further comprising performing human-face recognition using the resulting human-face global feature vectors.

8. The method according to claim 6, further comprising performing human-face recognition using the resulting human-face global feature vector set.

9. An apparatus comprising means for performing the actions of the method according to any of claims 1 to 8.

10. A computer program code configured to realize the actions of the method or any of claims 1 to 8, when executed by a processor.

Patentansprüche

1. Verfahren, das Folgendes umfasst:

Erhalten einer Videosequenz, die einen Satz von Gesichtsbildern enthält, die zur selben Person gehören;
Detektieren (S210) von Bildern des Gesichts in benachbarten Einzelbildern der Einzelbilder der Videosequenz;
Unterteilen (S220) der detektierten Bilder des Gesichts in mehrere lokale Bereiche mit einer vorgegebenen Größe, wobei jeder lokale Bereich einer Fläche des Gesichts um einen Merkmalspunkt des Gesichts entspricht;
Bestimmen (S230) einer Übereinstimmungsbeziehung zwischen den jeweiligen Merkmalspunkten des Gesichts der Bilder des Gesichts in den benachbarten Einzelbildern der mehreren Einzelbilder der Videosequenz;
Identifizieren von unterschiedlichen Stellungen des Gesichts in mehreren Einzelbildern einer Videosequenz;
Bestimmen, ob eine Fläche des Gesichts verdeckt ist, auf der Grundlage der identifizierten unterschiedlichen Stellungen des Gesichts in den mehreren Einzelbildern; **dadurch gekennzeichnet, dass** das Verfahren ferner Folgendes umfasst:

Extrahieren von Vektorkomponenten von lokalen Merkmalen des Gesichts nur aus lokalen Bereichen, die unverdeckten Flächen des Gesichts entsprechen, die eine wechselseitige Übereinstimmungsbeziehung aufweisen, und Verwerfen des Extrahierens von lokalen Merkmalen des Gesichts aus lokalen Bereichen, die verdeckten Flächen des Gesichts entsprechen, und
Kombinieren der Vektorkomponenten von lokalen Merkmalen des Gesichts aus loka-

len Bereichen in benachbarten Einzelbildern, die unverdeckten Flächen des Gesichts entsprechen, die eine wechselseitige Übereinstimmungsbeziehung aufweisen, um einen Vektor von lokalen Merkmalen des Gesichts zu bilden, der die Gesichtspunkte repräsentiert, die den lokalen Bereichen entsprechen.

2. Verfahren nach Anspruch 1, wobei das Unterteilen des detektierten Bildes des Gesichts in mehrere lokale Bereiche mit einer vorgegebenen Größe durch Hinzufügen eines Gitters zu dem detektierten Bild des Gesichts implementiert wird.

3. Verfahren nach Anspruch 1 oder 2, das ferner das Ändern der Größe der mehreren lokalen Bereiche umfasst, um die Vektoren der lokalen Merkmale des Gesichts zu erhalten, die die Gesichtspunkte repräsentieren, die den jeweiligen in der Größe veränderten lokalen Bereichen entsprechen.

4. Verfahren nach Anspruch 3, das ferner das Kombinieren der resultierenden Vektoren der lokalen Merkmale des Gesichts, die die Gesichtspunkte repräsentieren, die jeweiligen lokalen Bereichen entsprechen, umfasst, um Vektoren der globalen Merkmale des Gesichts zu bilden.

5. Verfahren nach Anspruch 4, das ferner das Kombinieren der resultierenden Vektoren der lokalen Merkmale des Gesichts, die die Gesichtspunkte repräsentieren, die jeweiligen in der Größe veränderten lokalen Bereichen entsprechen, umfasst, um Vektoren der globalen Merkmale des Gesichts zu bilden.

6. Verfahren nach Anspruch 5, das ferner das Kombinieren von Vektoren der globalen Merkmale des Gesichts, die unter unterschiedlichen Größen der lokalen Bereiche bezogen werden, umfasst, um einen Vektorsatz der globalen Merkmale des Gesichts zu beziehen.

7. Verfahren nach Anspruch 4 oder 5, das ferner das Durchführen einer Erkennung des Gesichts unter Verwendung der resultierenden Vektoren der globalen Merkmale des Gesichts umfasst.

8. Verfahren nach Anspruch 6, das ferner das Durchführen einer Erkennung des Gesichts unter Verwendung des resultierenden Vektorsatzes der globalen Merkmale des Gesichts umfasst.

9. Vorrichtung, die Mittel zum Durchführen der Schritte des Verfahrens nach einem der Ansprüche 1 bis 8 umfasst.

10. Computerprogrammcode, der konfiguriert ist, die Schritte des Verfahrens nach einem der Ansprüche 1 bis 8 zu verwirklichen, wenn es durch einen Prozessor ausgeführt wird.

Revendications

1. Procédé, comprenant :

l'obtention d'une séquence vidéo comprenant un ensemble d'images de visage appartenant à une même personne ;
la détection (S210) d'images de visage humain dans des trames adjacentes d'une pluralité de trames d'image de la séquence vidéo ;
la division (S220) des images de visage humain détectées en une pluralité de patchs locaux d'une taille prédéterminée, chaque patch local correspondant à une zone de visage humain autour d'un point de caractéristique de visage humain ;
la détermination (S230), au moyen d'un champ de flux optique, d'une relation de correspondance entre des points de caractéristique de visage humain respectifs des images de visage humain dans les trames adjacentes de la pluralité de trames d'image de la séquence vidéo ;
l'identification de poses différentes du visage humain dans une pluralité de trames d'image d'une séquence vidéo ;
la détermination si une quelconque zone de visage humain est masquée sur la base des poses différentes identifiées du visage humain dans la pluralité de trames d'image ; le procédé étant **caractérisé en ce qu'**il comprend en outre l'extraction de composantes de vecteur de caractéristiques locales de visage humain uniquement à partir de patchs locaux correspondant à des zones de visage humain non masquées présentant une relation de correspondance mutuelle, et l'omission de l'extraction de composantes de vecteur de caractéristiques locales de visage humain à partir de patchs locaux correspondant à des zones de visage humain masquées ; et la combinaison des composantes de vecteur de caractéristiques locales de visage humain extraites à partir de patchs locaux dans des trames adjacentes correspondant à des zones de visage humain non masquées présentant une relation de correspondance mutuelle de manière à former un vecteur de caractéristiques locales de visage humain représentant les points faciaux correspondant aux patchs locaux.

2. Procédé selon la revendication 1, dans lequel la division de l'image de visage humain détectée en une pluralité de patchs locaux d'une taille prédéterminée

est réalisée par ajout d'une grille sur l'image de visage humain détectée.

3. Procédé selon la revendication 1 ou 2, comprenant en outre le redimensionnement de la pluralité de patchs locaux de manière à obtenir les vecteurs de caractéristiques locales de visage humain représentant les points faciaux correspondant à des patchs locaux redimensionnés respectifs.

4. Procédé selon la revendication 3, comprenant en outre la combinaison des vecteurs de caractéristiques locales de visage humain résultants représentant les points faciaux correspondant à des patchs locaux respectifs de manière à former des vecteurs de caractéristiques globales de visage humain.

5. Procédé selon la revendication 4, comprenant en outre la combinaison des vecteurs de caractéristiques locales de visage humain résultants représentant les points faciaux correspondant à des patchs locaux redimensionnés respectifs de manière à former des vecteurs de caractéristiques globales de visage humain.

6. Procédé selon la revendication 5, comprenant en outre la combinaison de vecteurs de caractéristiques globales de visage humain obtenus selon des tailles différentes de patchs locaux de manière à former un ensemble de vecteurs de caractéristiques globales de visage humain.

7. Procédé selon la revendication 4 ou 5, comprenant en outre la réalisation d'une reconnaissance de visage humain au moyen des vecteurs de caractéristiques globales de visage humain résultants.

8. Procédé selon la revendication 6, comprenant en outre la réalisation d'une reconnaissance de visage humain au moyen de l'ensemble de vecteurs de caractéristiques globales de visage humain résultant.

9. Appareil, comprenant des moyens de réalisation des actions du procédé selon l'une quelconque des revendications 1 à 8.

10. Code de programme d'ordinateur configuré pour accomplir les actions du procédé de l'une quelconque des revendications 1 à 8 lorsqu'il est exécuté par un processeur.
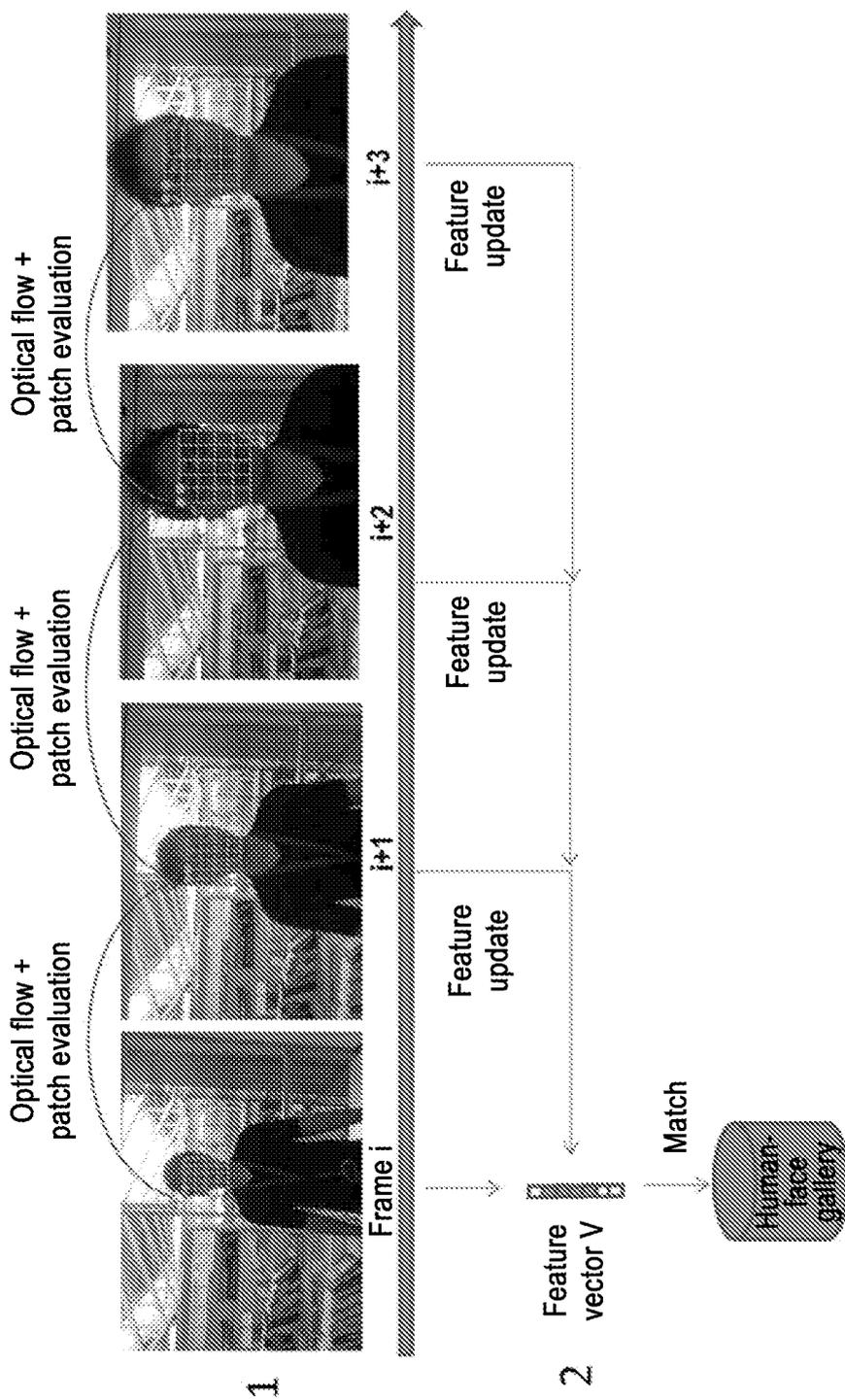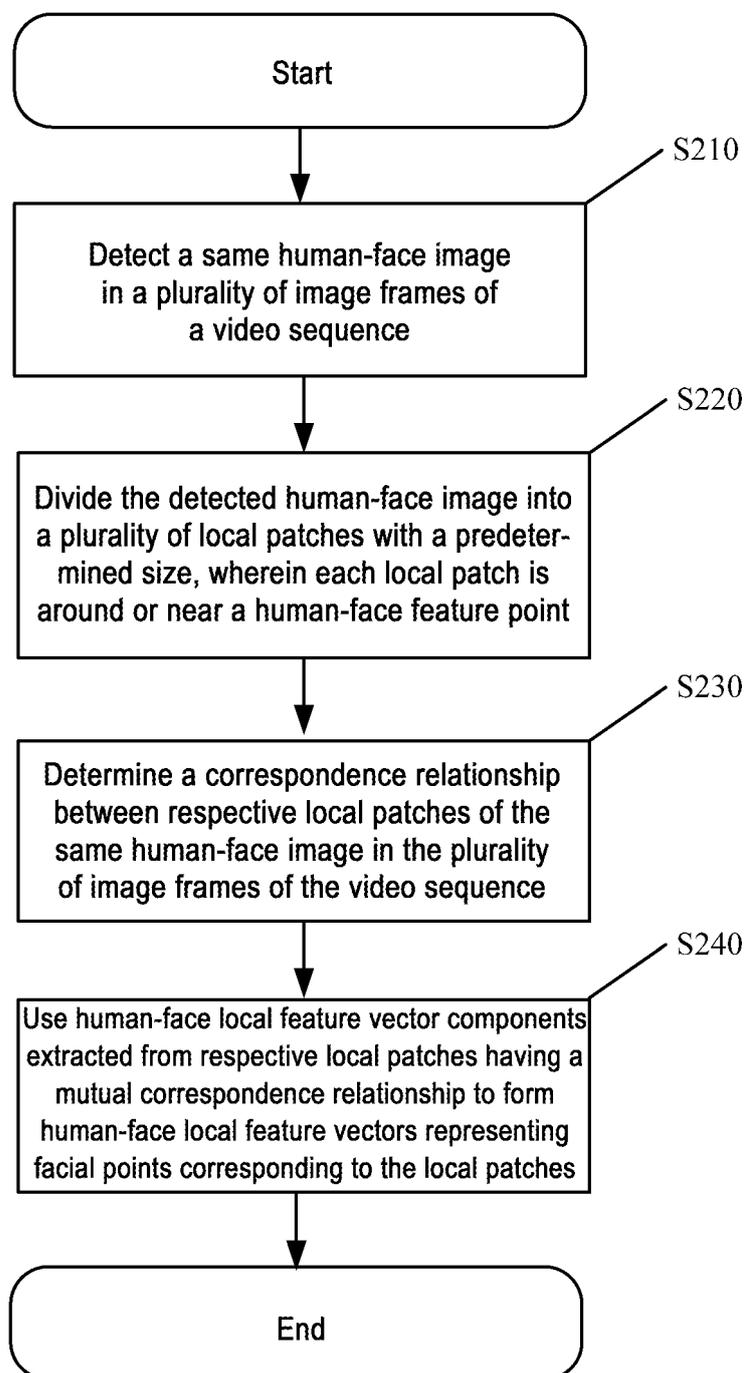
Fig. 1

Start

S210

Detect a same human-face image
in a plurality of image frames of
a video sequence

S220

Divide the detected human-face image into
a plurality of local patches with a predeter-
mined size, wherein each local patch is
around or near a human-face feature point

S230

Determine a correspondence relationship
between respective local patches of the
same human-face image in the plurality
of image frames of the video sequence

S240

Use human-face local feature vector components
extracted from respective local patches having a
mutual correspondence relationship to form
human-face local feature vectors representing
facial points corresponding to the local patches

End

Fig. 2

Learn patch
correspondence
relationship

Frame i

Frame i+1

Fig. 3

16×16
Grid patches

8×8
Grid patches

6×6
Grid patches

64×64
Normalized
face image

Fig. 4

Fig. 5

610 Means for detecting a same human-face image in a plurality of image frames of a video sequence

620 Means for dividing the detected human-face image into a plurality of local patches with a predetermined size, wherein each local patch is around or near a human-face feature point

630 Means for determining a correspondence relationship between respective local patches of the same human-face image in the plurality of image frames of the video sequence

640 Means for using human-face local feature vector components extracted from respective local patches having a mutual correspondence relationship to form human-face local feature vectors representing facial points corresponding to the local patches

600
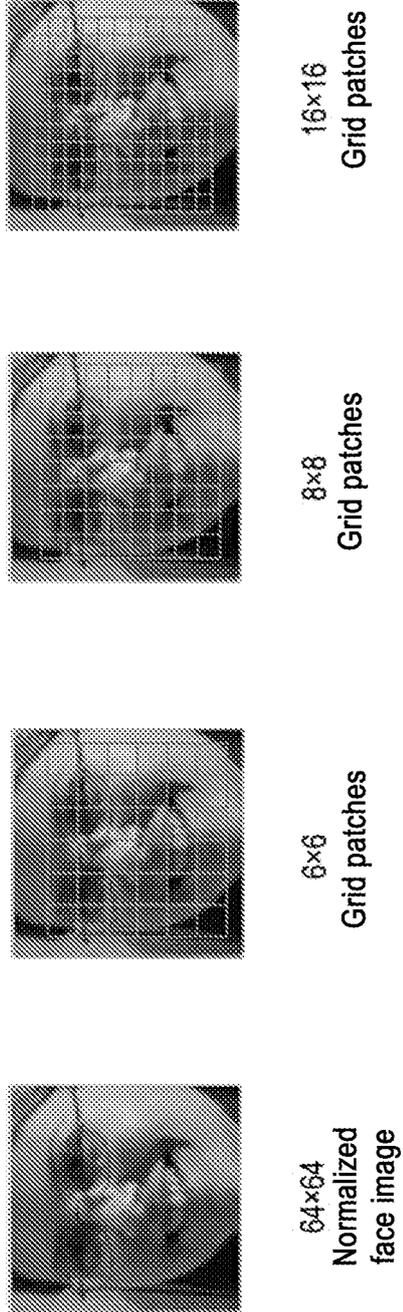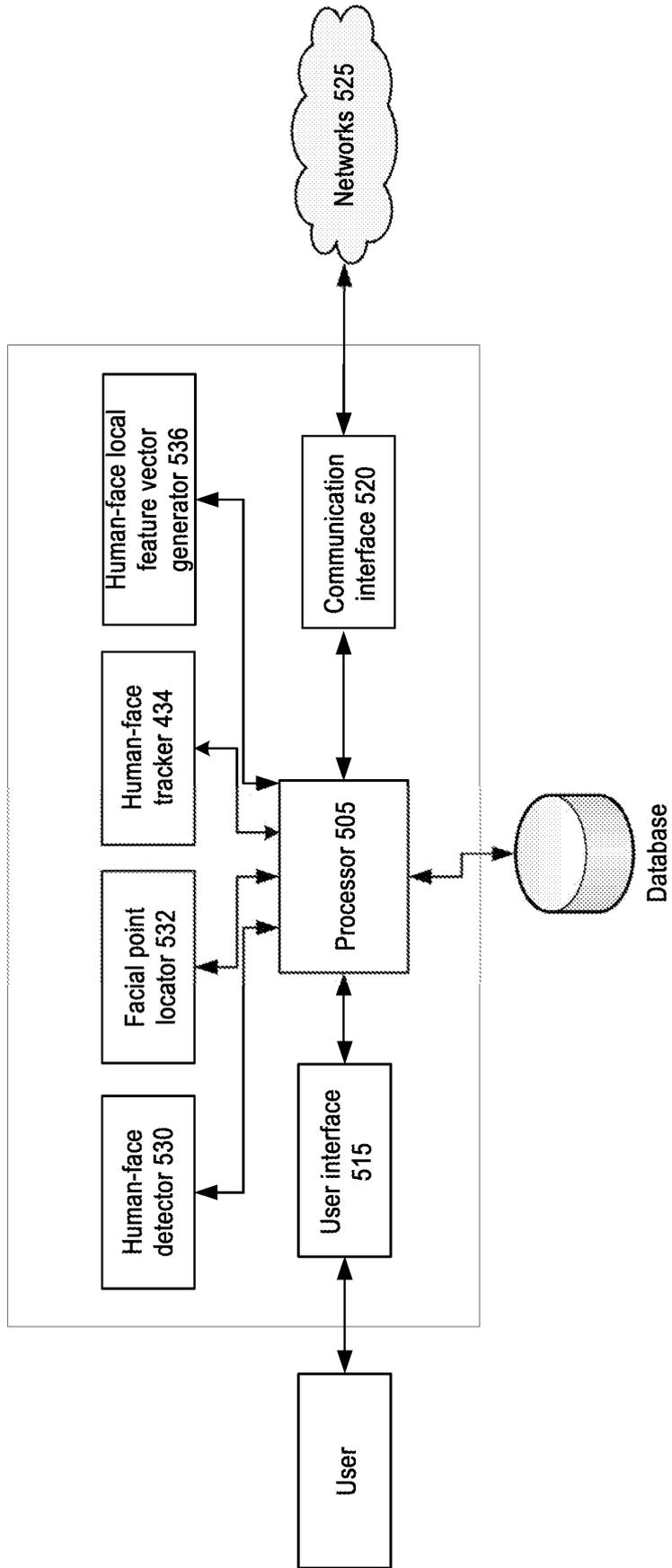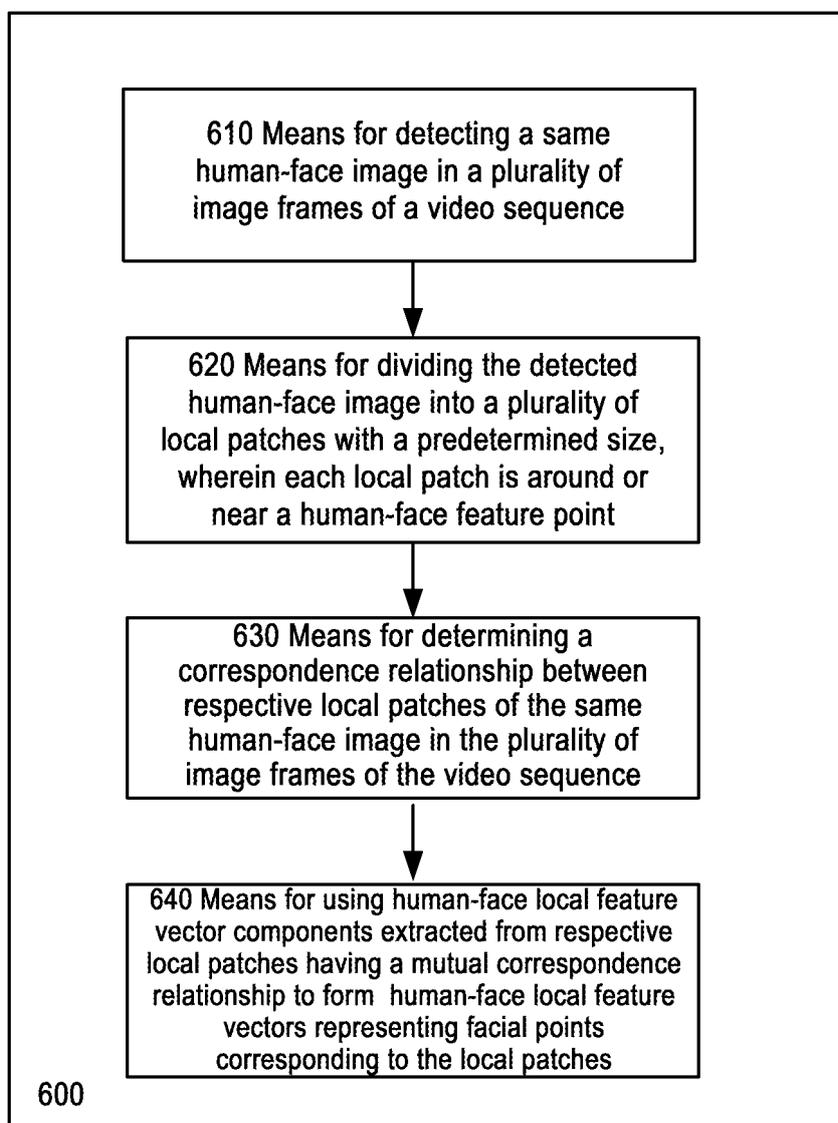
Fig. 6

REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Non-patent literature cited in the description

- **E. R GAST.** *A Framework for Real-Time Face and Facial Feature Tracking using Optical Flow Pre-estimation and Template Tracking,* April 2012 **[0012]**
- **X. LIU.** *Video-based face model fitting using Adaptive Active Appearance Model,* July 2010 **[0012]**
- Good features to track. **JIANBO SHI ; CARLO TOMASI.** Proc. IEEE Comput. Soc. Conf. Comput. Vision and Pattern Recogn, 1994, 593-600 **[0023]**
- **W. ZHAO ; R. CHELLAPPA ; A. KRISHNASWAMY.** Discriminant Analysis of Principal Components for Face Recognition. *Proc. of the 3rd IEEE International Conference on Face and Gesture Recognition, FG'98,* 14 April 1998, 336-341 **[0023]**
- **A.M. MARTINEZ ; A.C. KAK.** *PCA versus LDA, IEEE Trans. on Pattern Analysis and Machine Intelligence,* 2001, vol. 23 (2), 228-233 **[0023]**
- **OJALA ; MAENPAA.** Multiresolution gray-scale and rotation invariant texture classification with Local binary Patterns. *IEEE PAMI,* 2002, vol. 24 (7), 971-987 **[0028]**